
Articles

Calculating the Probability of Rare Events: Why Settle for an Approximation?

Harold S. Luft and Byron Wm. Brown, Jr.

Objective. Health services researchers often need to compute the probability of observing a certain number of events when only a few such events are expected. Our objective is to show that the standard approaches (Poisson, binomial, and normal approximations) are inappropriate in such instances, and to suggest an alternative.

Data Sources. Patients undergoing cholecystectomy (34,234) in 465 California hospitals in 1983 are used to demonstrate the biases arising from various methods of calculating the probability of observing a given number of deaths in each hospital. Similar data from other procedures and diagnoses with lower and higher mortality rates are also used for illustration.

Study Design. The computational methods to derive probabilities using the Poisson, normal, simulation, and exact probabilities are discussed. Using a previously developed risk factor model, the probability of observing the actual number of deaths (or more) is calculated given the expectation of death for each patient in each hospital. Results for the four methods are compared, showing the types of random and systematic errors in the Poisson, normal, and simulation approaches.

Data Collection. Routinely collected hospital discharge abstract data were provided by the California Office of Statewide Planning and Development.

Principal Findings. The Poisson and normal approximations are often biased substantially in calculating upper-tail p -values, especially when the expected number of adverse outcomes is less than five. Simulations allow unbiased calculations, and the degree of random error can be made arbitrarily small given enough trials. Exact calculations using a simple recursive algorithm can be done very efficiently on either a mainframe or personal computer. For example, the whole set of cholecystectomy patients can be assessed in less than 90 seconds on a Macintosh®.¹

Conclusions. Calculating the probability of observing a small number of events using standard approaches may result in substantial errors. The availability of a simple and inexpensive method of calculating these probabilities exactly can avoid these errors.

Keywords. Statistics, small-area analysis, risk adjustments, computational techniques

Investigators in the health services research setting frequently observe a particular outcome in a group of patients and wish to calculate the probability that that outcome occurs by chance. For example, a large data base may be used to determine the risk factors associated with patient mortality for patients undergoing a particular procedure. These risk factors are then applied to the mix of patients in specific hospitals and the expected number of deaths is compared with the observed number. Hospitals with significantly higher than expected death rates may then be examined for potential quality of care problems. Alternatively, one may be examining the number of asthma deaths across geographic areas and wish to know whether the numbers observed in certain counties are excessive (Weiss and Wagener 1990). Likewise, small-area admission rates are usually calculated relative to a demographically adjusted expected number. In studies like these, there are usually large numbers of observed and expected cases in each hospital, small geographic area, or similar unit of observation. However, it is not uncommon that some of the smaller "sites" will have fewer than five expected cases.

The usual approach in such situations is to calculate the probability based upon the Poisson distribution or the normal approximation to the binomial. For example, Weiss and Wagener (1990) used a chi-square test if the number of observed deaths was 30 or more; otherwise, the probability was computed using a Poisson distribution with mean equal to the expected number of deaths. Escarce and Kelley (1990) used a chi-square test when both observed deaths and expected deaths exceeded five; otherwise, tests based on the binomial distribution were used.

The usual rule of thumb offered in statistics texts is that the normal distribution is a reasonably good approximation if the number of expected outcomes (e.g., deaths) is five or more (Fleiss 1981). How-

This research was supported by grants HS06216, "Volume, Outcome, Market Shares, and Admission Rates," and HS06503, "Outcome Assessment Program in Ischemic Heart Disease," from the Agency for Health Care Policy and Research (AHCPR). Data were provided by the California Office of Statewide Health Planning and Development.

Address correspondence and requests for reprints to Harold S. Luft, Ph.D., Professor and Acting Director, Institute for Health Policy Studies, School of Medicine, University of California, 1388 Sutter Street, San Francisco, CA 94109. Byron Wm. Brown, Jr., Ph.D. is Professor and Chair of the Department of Health Research and Policy, Stanford University School of Medicine. This article, submitted to *Health Services Research* on July 25, 1991, was revised and accepted for publication on November 30, 1992.

ever, there is generally little discussion of the risks associated with using the normal when the expected number of cases is small, and it is often not recognized that the Poisson is also an approximation with a similar rule of thumb. In particular, the Poisson and normal distributions may provide poor approximations of the true probabilities when the numbers of the observed and expected deaths are small. This may lead to the incorrect interpretation of some situations as being "significant statistically" when they are not, and contrariwise, to the failure to detect statistically significant aberrations. This article provides examples of these problems in "real world" health services research situations and a practical alternative for the calculation of probabilities that is feasible on personal or mainframe computers even for large-scale analyses, that is, many study units with many subjects per unit.

DATA AND METHODS

To illustrate the problem and the proposed solutions, we use hospital discharge abstract data obtained from the California Office of Statewide Health Planning and Development for patients discharged in 1983 from California hospitals. We first provide examples of the problems associated with various calculation approaches using patients undergoing cholecystectomy. We then test implications of various approaches on a series of diagnoses and procedures with various death rates and numbers of patients per hospital (see Table 1). The overall mortality rate for cholecystectomy patients in our data is 1.16 percent, which is approximately in the middle of the range among the patient groups studied; hysterectomy has a death rate of 0.05 percent, and acute myocardial infarction is at the other extreme with a 15.0 percent death rate. The average number of cholecystectomy patients per hospital is 74. By contrast, average volumes range from 244 coronary artery bypass graft (CABG) patients to 21 hip replacement patients per hospital.

The risk of death is usually not the same for all patients, so it is important to compute an expected probability of death for each patient. In each instance, we excluded patients for whom the procedure seemed not to be the reason for admission to the hospital, for example, cholecystectomy in response to abdominal trauma. A logistic regression was estimated for each patient group using inpatient death or survival as the dependent variable and age, gender, chronic comorbidities, emergency status, and selected related procedures as the independent variables. By applying the estimated coefficients from the logistic regression to each patient's characteristics, we obtain an estimated

Table 1: Patients, Deaths, and Expected Deaths for Selected Patient Groups, California Hospitals, 1983

	Total									
	Number		Death		Patients/Hospital		Deaths/Hospital		Expected Deaths/Hospital†	
	Hospitals	Patients	Rate*	Rate*	Minimum	Maximum	Average	Maximum§	Minimum	Maximum
Cholecystectomy	465	34,234	.0116		2	445	0.85	8	0.0008	5.33
Pneumonia	484	47,095	.1069		2	524	10.40	73	0.0756	38.02
AMI‡	470	50,036	.1504		2	685	16.01	74	0.1038	142.21
CABG‡	78	19,064	.0354		20	939	8.65	31	0.3880	29.06
Femur fracture	442	24,721	.0351		2	412	1.96	14	0.0142	10.03
Hip replacement	320	6,839	.0050		1	192	0.11	1	0.0012	0.78
Hysterectomy	461	48,211	.0005		1	680	0.05	2	0.0000	0.63
Stroke	474	53,031	.1197		2	572	13.39	59	0.2600	60.00
TURP‡	426	22,512	.0028		1	475	0.15	3	0.0003	1.43

*Death rate = Total deaths/total patients.

†Expected deaths is calculated by summing the probability of death for each patient in the hospital. See text.

‡AMI = acute myocardial infarction; CABG = coronary artery bypass graft; TURP = transurethral resection of the prostate.

§Minimum number of deaths observed is always zero.

probability of death. (The particular variables used and the coefficients of the logistic regressions are not crucial for this discussion. They may be obtained from the authors.) We then produce a file with three data items for each patient: (1) the index j for the hospital in which the patient was treated, (2) the estimated probability of death ($p_{i,j}$) for patient i in hospital j , and (3) whether the patient actually died ($d_{i,j}$) where 1 = death, 0 = survival. This patient-level file can then be used to calculate statistics at the hospital level by aggregating patients in each hospital.

The usual problem is that some number of deaths is observed at a hospital $d^*_j = \sum_i d_{i,j}$ when $p^*_j = \sum_i p_{i,j}$ were expected, so the task at hand is to calculate the probability of this event or one more extreme (i.e., more deaths) occurring by chance. In this example the probability of death will vary from patient to patient. If there are n_j patients in the j th hospital, this is formally a mixture of n_j Bernoulli distributions, and the total number of deaths d^*_j among the n_j patients is said to have a Lexis distribution (Murphy 1979). The Lexis distribution is rarely used explicitly for analytic purposes; it depends on the individual probabilities of death and creates computational problems if used to obtain exact results.

There are several approaches to the calculation of the probability of observing d^*_j or more deaths in the j th hospital: the determination of the probability by an exact calculation, approximations based on the Poisson and normal distributions, and simulations. In each case we will assume that the estimated probabilities of death for each patient are known without error. While this is not generally the case, the logistic regressions are based on far more patients than occur in any single hospital. (For example, there are 34,234 cholecystectomy patients in our data file; the most in any one hospital is 445.) Thus, the chance errors associated with the estimation of the patient-level probabilities are small relative to the role of chance in the observed outcomes among the patient population of any given hospital. This is not to imply, however, that large numbers assure the inclusion of the correct risk factors, but that given the state of knowledge about the relevant risk factors, the underlying probabilities are assumed to be known. Note that this is the standard approach in the literature (Weiss and Wagener 1990; Escarce and Kelley 1990; Flood, Scott, Ewy, et al. 1987).

To simplify the following discussion, we will focus on only patients in a given hospital, so we can drop the j subscript. The probability of d^* or more deaths, where d^* is small, is more easily calculated by first computing its complement, the probability of observing fewer than

d^* deaths. Likewise, the probability of survival for a given patient, q_i , is $1 - p_i$. If there are two patients, the probability that both will survive is $q_1 * q_2$. If there are three patients, the probability that all will survive is $q_1 * q_2 * q_3$. One minus these products is the likelihood of one or more deaths among these patients. For two or more deaths, one first calculates the probability of no deaths, as before, plus all of the ways in which one death could occur. The latter is the probability that the first patient dies, p_1 , times the probability that all of the others survive, plus the probability that the second patient dies times the probability that all of the others survive, and so forth. A more complete discussion of the exact calculations is in an appendix available from the authors.

The exact probability calculations are quite straightforward and easily coded in programming languages such as FORTRAN or C. The reason they are not routinely implemented in most statistical packages is because of the computational costs incurred using the most obvious programming approaches. This cost is roughly proportional to the number of patients raised to the d^* power. Thus, for 100 patients, the computer time required to compute the probability of five or more deaths is 100^4 , or 100,000,000 times that required to compute the probability of one or more deaths.

For large numbers one can rely on the fact that well-known distributions approximate the exact probabilities. For example, when each patient either lives or dies with a known probability, the probability that d^* or more deaths is observed is approximately normal, provided the expected number of deaths is not too small. (The general rule of thumb for "too small" is fewer than five.) A two-step process can then be used to calculate the probability of observing d^* or more deaths. First, one computes the standardized deviate (the Z-score), which is essentially the number of standard deviations the observed value is away from the mean or expected value. (One usually subtracts 1/2 from the difference between the observed value and the mean to adjust for the fact that deaths occur in integral amounts, but the normal is continuous. This is called the continuity correction.) Since the probability of death for each patient, p_i , is assumed to be known, it is straightforward to compute both the total number of deaths expected and the standard deviation, which is the square root of $\sum p_i q_i$. (Note that if the probability of death for all patients were the same, say p^* , then the standard deviation is the square root of $np^*(1 - p^*)$, the familiar binomial expression.) The second step is the conversion of the Z-score to a one-tailed probability. The standard normal tables can be used for this purpose, or there are quite precise simple formulas that can be

incorporated in a computer program. Most statistical packages include such a program for normal tail areas.

When events occur very infrequently, as is sometimes the case for deaths in certain diagnostic or procedure groupings, the number of observed deaths, d^* , may be approximately Poisson distributed. With the Poisson distribution, the calculation of the probability of observing d^* or more deaths depends only on the expected number of deaths, $s = \sum p_i$. The formula is $\text{Prob}\{d \geq d^*\} = 1 - \{e^{-s} + e^{-s}s + e^{-s}s^2/2! + \dots + e^{-s}s^{d^*-1}/(d^*-1)!\}$. The calculation of the probabilities is simple and fast, and the approximation will be satisfactory if the p_i are all small and fairly uniform but the n is large enough so that $s \geq 5$. Again, standard statistical packages usually have a component available to calculate tail probabilities based on the Poisson distribution.

A fourth approach to calculating the probability of observing a given number of deaths is simulation. This allows one to estimate the actual distribution of outcomes likely in a set of patients, each of whom has a specific probability of dying. The approach is straightforward. A number is drawn from a uniform random distribution with a range from 0 to 1. If the number drawn is less than the patient's probability of dying, the patient is counted as dead, otherwise, alive. A new number is drawn for each patient in the hospital, thus obtaining an "observed" number of deaths given a single set of n random draws. If one then repeats this process K times, one can then count the number of times (i.e., simulations of the n patients) for which d^* or more deaths are observed. That number divided by K is the estimated probability of observing that many or more deaths.

Unlike the exact calculation of p and the normal and Poisson approximations, which will always provide the same result given the same set of patient probabilities, simulations introduce random variability *into the calculation*, quite apart from the variability in outcomes that one is trying to examine. This variability can be of some concern in situations involving low-probability events. For example, if the true tail probability for a given outcome (d^*) is 1 percent, there is a 5 percent chance that a simulation with 1,000 trials could yield an estimate as low as 0.4 percent or as high as 1.6 percent. These limits of precision can be substantially reduced if the number of simulations is increased. The advantage of the simulation (or Monte Carlo) method is that it provides an unbiased estimate of the desired probability (if the random number generator is valid), with precision depending only on the number of simulations. The disadvantage is the computing time or cost since the desired precision may require a great number of simulations when the true tail probability is small, as in our applications.

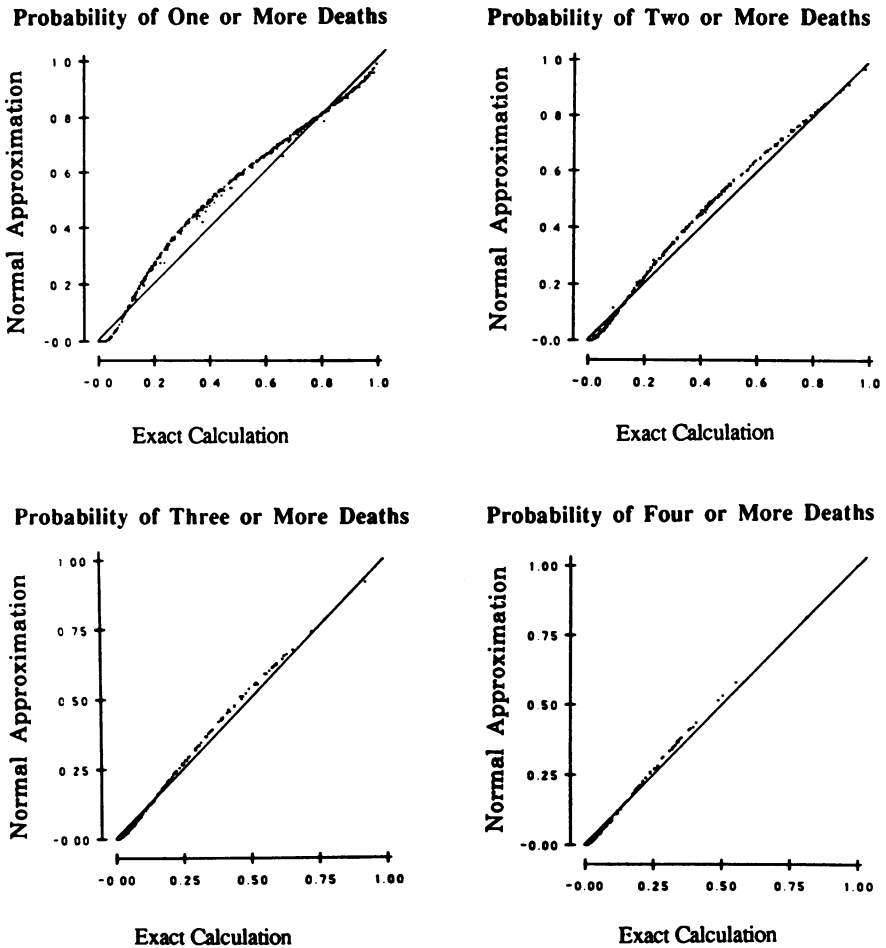
COMPARISON OF ALTERNATIVE MEASURES

The set of cholecystectomy patients provides an example of applying the various approaches. The number of patients in the 465 hospitals ranges from 2 to 445, with the total across all hospitals being 34,234. The expected number of deaths ranges from 0.0008 to 5.3265. The number of actual deaths ranges from 0 to 8. The general rule of thumb is that if five or more deaths are expected, then the normal approximation is reasonable to use. The standard literature, however, does not offer a simple rule of thumb for gauging the error that might be introduced if the expected number of deaths is smaller than this, or if the $\sum p_i q_i$ departs much from $np^*(1 - p^*)$, where p^* is the average death rate for all patients within a particular hospital.

In this section we explore the implications of using different approaches to approximating the probability associated with a given number of deaths. In doing so, we use the estimated probability of death for each individual patient, but for the moment ignore the actual number of deaths observed in each hospital. The "gold standard" will be the exact calculation of the probability associated with four situations: observing at least d^* deaths among the patients in each hospital, where d^* ranges from 1 to 4.

Figure 1 presents the probability of observing d^* or more deaths among the cholecystectomy patients in each of the 465 hospitals in the data set. These graphs present the probability calculated using the normal approximation on the vertical axis and the exact probability on the horizontal axis. (The standard deviation for each normal deviate, Z , is based on $\sum p_i q_i$ for that hospital rather than the binomial expression, $np^*(1 - p^*)$.) Although the normal approximation seems reasonably good, the curves indicate some important deviations. First, there is a "bowing" out of the curve—most noticeable for the probability of one or more deaths but nonetheless detectable for the other outcomes as well. That is, over much of the range, the normal approximation tends to overestimate the true p -value associated with the number of observed deaths relative to expected deaths. At very low probabilities, the normal approximation seems to "bottom out" at zero. Or, putting it another way, it reaches zero too soon, underestimating the probability at the low end of the range. Likewise, it underestimates probabilities at the high end. While there appears to be a fairly consistent pattern of over- or underestimation that depends on the value of the true proba-

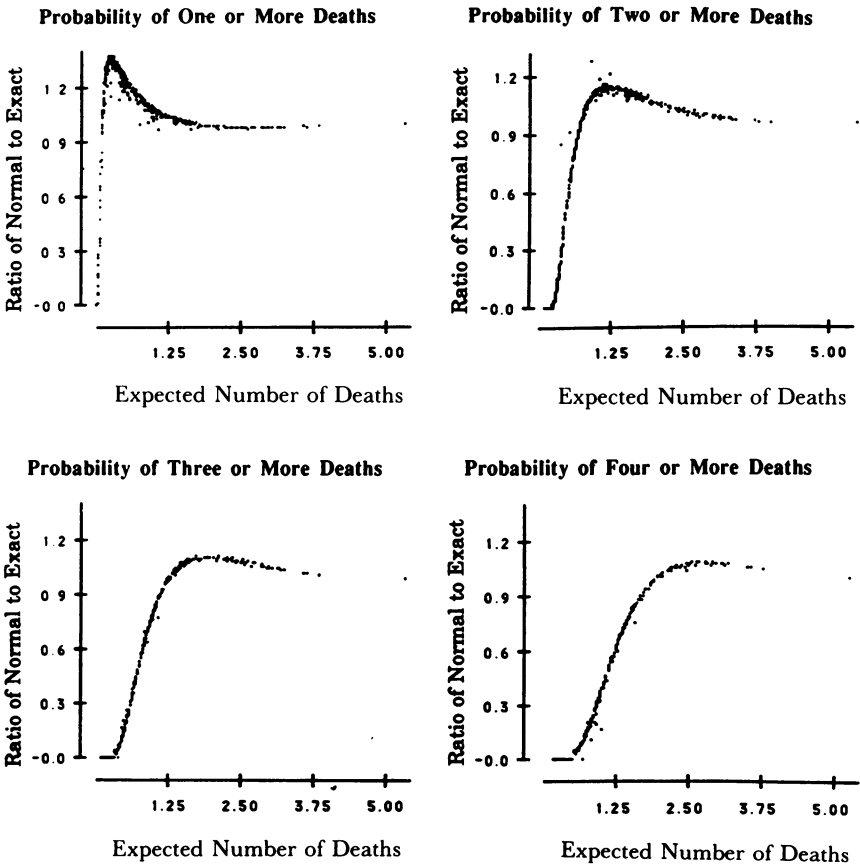
Figure 1: Comparison of Probabilities Calculated Using Normal Approximation and Exact Calculations Based on Cholecystectomy Patients in 465 California Hospitals



bility, there are occasional hospitals that deviate substantially from this pattern.

Figure 2 highlights the errors more clearly by plotting the ratio of the normal approximation to the exact probability as a function of the expected number of deaths in the hospital. All three problems with the normal approximation are now quite apparent. When the number of expected deaths is very small, the normal approximation-based proba-

Figure 2: Ratio of Probabilities Using Normal and Exact Calculations versus Expected Number of Deaths in the Hospital Based on Cholecystectomy Patients in 465 California Hospitals



bility underestimates the true value. This problem is more extreme close to zero. As the expected number of deaths in a “sampling unit” (hospital) rises, the probability based on the normal approximation rises rapidly relative to the true probability and eventually exceeds it by up to 30 percent. Then, as the expected number of deaths increases further, the ratio falls to asymptotically approach 1. For larger num-

bers of observed deaths, the "safe range" for the normal approximation to the exact values moves out to larger numbers of expected deaths. With one or more observed deaths, expected values of 2.5 or more seem to yield good agreement. For the probability of four or more deaths, an expected number of deaths of 4 might be required to feel comfortable using the normal approximation.

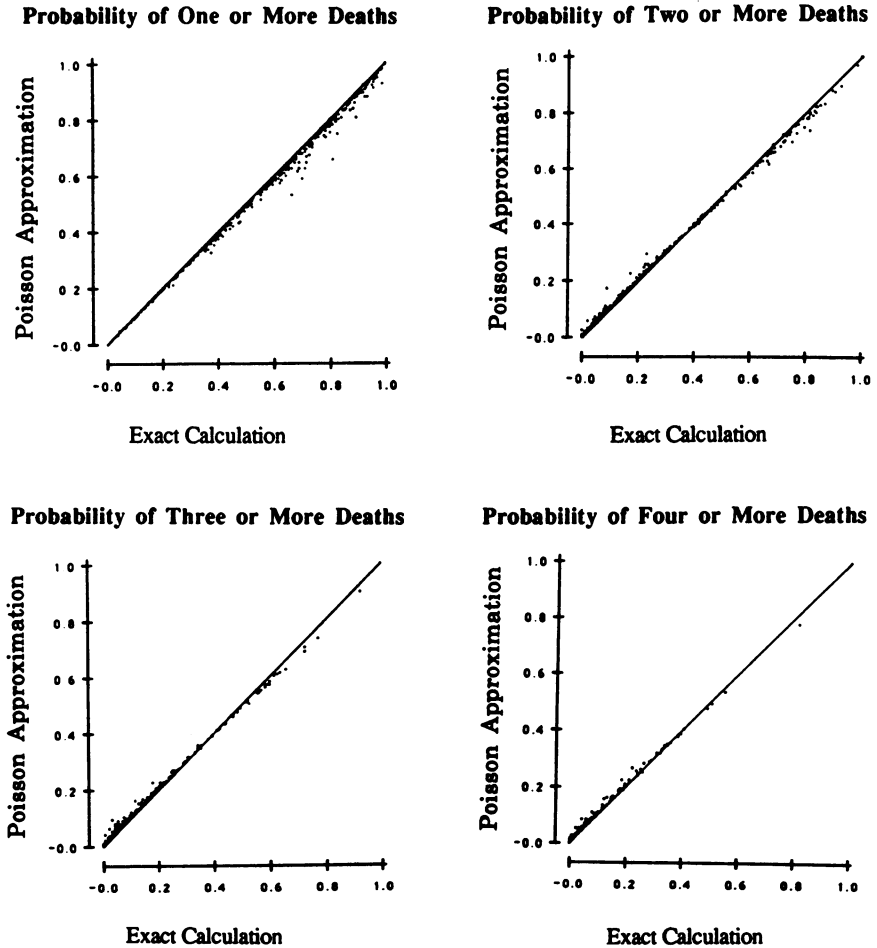
Figure 3 is analogous to Figure 1, except that the probabilities on the vertical axis were calculated using the Poisson distribution. Again, the Poisson approximation is fairly good, but with problems also apparent. For the probability of one or more deaths, a substantial number of instances occur in which the Poisson approximation substantially underestimates the true probability, particularly when the probabilities are large. For two or more deaths, the Poisson outliers tend to be overestimates when the probabilities are low and underestimates when they are large.

The ratios of Poisson to exact probabilities are plotted in Figure 4. Here it is apparent that in some ways the situation is worse than was the case for the normal approximation. With relatively small numbers of expected deaths, the normal probabilities deviate markedly from the exact probabilities, but the pattern of over- and underestimation is quite predictable, and one might hope for a transformation of the normal figures to the correct values. The Poisson probabilities, on the other hand, are usually underestimates of the true values for $d^* = 1$ and overestimates for the other three cases. The percentage errors can be quite large and, more importantly, they do not follow a consistent pattern as is the case for the normal probabilities. This should not be surprising, since the Poisson depends only on the expected number of deaths, and thus ignores the variability among patients within a hospital.

Similar figures were developed for probabilities based on 1,000 and 10,000 simulations. While there is substantial scatter with 1,000 simulations, the approach assures the lack of bias, in contrast to the normal and Poisson approximations. With 10,000 simulations for each patient the scattering around the equal probability line is quite small. In percentage terms, however, even these errors may be substantial for very low probability events.

One way of assessing the problems associated with the various methods of calculating probabilities is to see how often they would lead one to misinterpret a statistically significant event as being insignificant, and vice versa. While comparison with an arbitrary p -value is not a recommended decision rule, it is a common practice. Thus, we can ask, if the true probability of an event is within a certain range, how

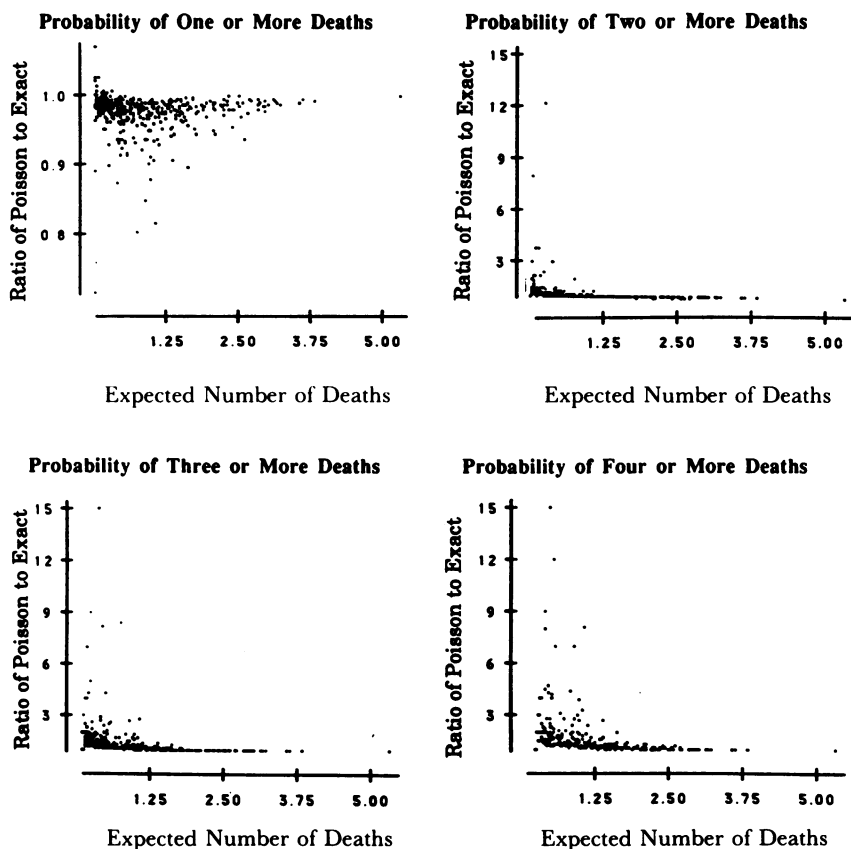
Figure 3: Comparison of Probabilities Calculated Using Poisson Approximation and Exact Calculations Based on Cholecystectomy Patients in 465 California Hospitals



often will the calculation based on the various alternatives be within the same range?

In examining four possible outcomes for each of the 465 hospitals (i.e., $d^* = 1, 2, 3$, and 4), we have a total of 1,860 comparisons for each method of calculation. Five categories are used, $p < .001$, $.001 \leq p < .01$, $.01 \leq p < .05$, $.05 \leq p < .10$, and $.10 \leq p$. Table 2 presents the percentage of times the normal, Poisson, and simulation approaches

Figure 4: Comparison of Probabilities Using Poisson and Exact Calculations versus Expected Numbers of Deaths in the Hospital Based on Cholecystectomy Patients in 465 California Hospitals



The vertical scales differ between panel 1 and the other panels. In panel 4 one point not shown has a value of 75.

(with 1,000 and 10,000 trials) give various approximate values for the true p . The normal approximation gives probabilities in the correct range almost all the time for p -values of less than .001 or greater than .10. For the range when the tests are more likely to matter, .001 to .10,

Table 2: Distribution of Probabilities Calculated Using Various Methods Given an Exact Probability, Based on 1, 2, 3, and 4 or More Deaths among Cholecystectomy Patients in California Hospitals

	True Probability				
	$p < .001$	$.001 \leq p < .01$	$.01 \leq p < .05$	$.05 \leq p < .10$	$.10 \leq p$
<i>Probabilities based on normal approximation</i>					
$p < .001$	100.0%	84.6%	7.4%	—	—
$.001 \leq p < .01$	—	15.4	47.9	—	—
$.01 \leq p < .05$	—	—	44.7	38.5%	—
$.05 \leq p < .10$	—	—	—	58.0	0.8%
$.10 \leq p$	—	—	—	3.6	99.2
<i>Probabilities based on Poisson approximation</i>					
$p < .001$	100.0%	21.7%	—	—	—
$.001 \leq p < .01$	—	78.3	20.3%	—	—
$.01 \leq p < .05$	—	—	79.7	17.6%	—
$.05 \leq p < .10$	—	—	—	82.4	2.2%
$.10 \leq p$	—	—	—	—	97.8
<i>Probabilities based on 1,000 simulations</i>					
$p < .001$	85.6%	8.9%	—	—	—
$.001 \leq p < .01$	14.4	82.1	5.4%	—	—
$.01 \leq p < .05$	—	8.9	88.7	7.1%	—
$.05 \leq p < .10$	—	—	5.8	84.6	0.2%
$.10 \leq p$	—	—	—	8.3	99.8
<i>Probabilities based on 10,000 simulations</i>					
$p < .001$	96.7%	3.7%	—	—	—
$.001 \leq p < .01$	3.3	94.3	1.9%	—	—
$.01 \leq p < .05$	—	2.0	96.5	2.4%	—
$.05 \leq p < .10$	—	—	1.6	95.9	0.2%
$.10 \leq p$	—	—	—	1.8	99.8

the normal approximation is biased downward. For example, 84.6 percent of the times when the true tail probability is between .001 and .01, the normal would place it below .001. When the true value is between .05 and .10, in 38.5 percent of the cases the normal would place it below .05 and in 3.6 percent of the cases the normal would place it above .10. Thus, probabilities based on the normal approximation would often mislead an investigator to believe the findings to be more significant than they were.

The Poisson has a similar problem as it often results in a calculated tail probability below the true value. For example, when the true

p -value is between .001 and .01 or between .01 and .05, the Poisson calculation results in a probability in the next-lower category more than 20 percent of the time. In the borderline range of .05 to .10, the Poisson approximation yields a "significant result" (i.e., $p < .05$) 17.6 percent of the time.

The unbiasedness of the simulation approach can be seen in the lower part of Table 2. Except for the extreme categories, over- and underestimates occur roughly equal numbers of times. With 1,000 simulations the errors are in the range of 11.3 percent to 17.9 percent, while with 10,000 simulations the errors range from 3.5 percent to 5.7 percent. Although the calculations based on 1,000 simulations fall in the correct category only somewhat more frequently than the Poisson estimates, they are unbiased. More importantly, the probabilities based on 10,000 simulations are always superior to the other three approaches, although obviously still not as good as exact calculations.

BENEFITS AND COSTS OF THE VARIOUS CALCULATION APPROACHES

These analyses of the cholecystectomy data lead to the following observations. For very small numbers of expected deaths, all three alternatives to an exact calculation yield rough approximations, that is, small probabilities; but these may differ substantially in percentage terms from the exact values. The simulation-based probabilities are unbiased and provide precise estimates even for patient groups with highly skewed probabilities of death, situations which result in particularly poor normal and Poisson approximations. With 10,000 simulations, the Monte Carlo errors appear to be tolerable.

In comparing the Poisson and normal approximation approaches, the latter seem preferable. They seem equally good in the range of three or more expected deaths. At low values of expected deaths the normal approximation probabilities exhibit a fairly consistent pattern of errors. This systematic bias might be modeled to provide a correction function, although we have not attempted to do so. The errors in the Poisson estimates, however, seem to fit no such pattern.

As indicated earlier, it is possible to compute exactly the probability of observing any number of deaths given the individual patient probabilities. Using standard computational approaches, cost is roughly proportional to $n^k/k!$ where n is the number of patients in a

given hospital or sampling unit and k is the number of observed deaths. An iterative method of generating the probabilities, however, is much simpler and its cost is roughly proportional to the number of patients times the number of deaths observed. The Z-score and probability calculation for the normal approximation is quite simple and its cost is roughly proportional to the number of patients. Simulations, on the other hand, are relatively expensive, because a large number of random numbers, for example, 10,000, have to be generated for each patient, but the cost is linear with respect to patients, and the simulations provide the frequency counts for all values of d^* .

The exact costs associated with any of these calculations depends on the hardware and software used. In this instance relatively simple and low-cost technology was used, a Macintosh® II with an internal Rodime hard drive, a 68881 floating point math coprocessor and 5 megabytes of memory, although less than 60K was needed. All programs were written in Microsoft®² FORTRAN, including a uniform random number generator following the model given in Knuth (1981). (Copies of the programs are available from the first author.)

Both the standard rule of thumb and our examination of the normal approximation results suggest that the normal approximation is reasonable when the number of expected deaths is five or more. This allows a dual approach—exact calculation if the number of expected deaths is less than five and the normal approximation otherwise. In fact, the exact probability code is so inexpensive we use it for all hospitals with 15 or fewer *observed* deaths. For the 465 hospitals with 34,234 patients, these calculations took less than 90 seconds. In contrast, 10,000 simulations for the same group of patients took about 2.25 hours. (Using a SAS®³ version on the mainframe, the exact computations took under five seconds.)

To help put this approach in perspective, we examined a pooled data set for all California hospitals for the period 1983–1986 for the set of nine procedures and diagnoses listed in Table 1. By pooling the data across four years we have a better chance of replicating the type of variability in outcomes and expected values across hospitals. Only 6 percent of the 1,828 hospital-year observations with cholecystectomy patients would have probabilities calculated using the normal approximation. In contrast, 83 percent to 87 percent of all hospital-years could use the normal approximation for pneumonia, stroke, acute myocardial infarction (AMI), and CABG patients. For total hip replacement, hysterectomy, and prostatectomy patients exact calculations would always be done. Even though hospitals average 2.09 deaths among

femur fracture patients (Table 1), 28 percent of the hospitals have five or more expected deaths, so a normal approximation could be used.

CONCLUSIONS

While this article focused on the methods to be used for calculating probabilities at the hospital level, the implications are broader. For example, measuring outcomes or admission rates across small geographic units raises comparable problems in determining whether a particular observed result is notably different from expected. A reliance on just the number of patients in the denominator is not sufficient. Likewise, requiring that there be a minimum number of observed "adverse" outcomes in the numerator is also inappropriate. In fact, the *expected* number of outcomes is the crucial variable in deciding whether a formula approach such as the normal approximation is warranted.

If fewer than five occurrences are expected, then the use of either the Poisson or normal approximation may result in inaccurate probability estimates. Our analyses suggest that these errors are relatively small in absolute magnitude, but may be large in relative terms. For example, in one instance the estimated probability based on the normal approximation was .0032 when the exact probability was .0298, nearly an order of magnitude error. If one focuses on the absolute levels, then both figures are seen to be fairly small. If one uses arbitrary cutoff levels, however, such as .01 for a "significant outlier," then such errors may be more harmful. Furthermore, if one is adjusting for multiple comparisons, then the "cutoff point" often becomes quite small. With the normal approximation, over a third of the instances in which hospitals had a true p -value between .05 and .10 would have been miscalled as being less than .05. For the Poisson, 17.6 percent of the instances in which the true p -value was between .05 and .10 would have been miscalled as less than .05.

Empirical analysis is difficult enough, and various errors in measurement are always a threat to the validity of a study. Thus, it is unwise to introduce additional errors into the analysis by using inappropriate approaches to calculating probabilities. Performing exact calculations of probabilities is a simple task, although it requires some programming rather than a simple request to a packaged data analysis system. A stand-alone program in either FORTRAN or SAS® requires about two pages of code, including extensive comments (Appendix A

and B). Moreover, the cost of exact calculations is far from prohibitive, even on a microcomputer, let alone a mainframe.

Recognizing that the use of normal or Poisson-based probabilities may introduce potentially important errors and that the correct approach is relatively simple leads us to suggest that careful attention be paid to how one computes probabilities. For those who find it necessary to compute probabilities in situations in which the expected number of outcomes is more than five, the conventional approaches are quite adequate. However, if the expected number of outcomes is frequently less than five, and if many calculations are required, the inclusion of specific programs to do exact calculations should not be a major hurdle.

APPENDIX A

A Method for Rapid Calculation of Exact Probabilities

The direct approach to the calculation of the exact probability of observing d^* or more deaths among a set of n patients with known probabilities of death p_i , $i = 1, 2, \dots, n$ requires the computation of all the possible combinations of $1, 2, \dots, (d^* - 1)$ deaths drawn from n patients. This computational effort is a function of $n^{(d^*-1)}$, which for large n and d^* of 4 or more is substantial.

A far more efficient approach uses intermediate sums in the computational process, so the number of calculations is roughly $2^*(d^* - 1)*n$. The method uses a series of d^* cumulative sums and is best understood using an array. The first two columns merely represent the probability of death for the i th patient, p_i , and the probability of survival, $1 - p_i$ or q_i . Subsequent columns represent the probability of observing 0, 1, 2, \dots , $d^* - 1$ deaths.

For the first patient, the entries are straightforward: q_1 is the probability of no deaths and p_1 the probability of one death. If a second patient is added, the probability of 0 deaths is $q_1 * q_2$; the probability of exactly one death is $q_1 p_2 + p_1 q_2$; the probability of two deaths is $p_1 p_2$. If a third patient is involved, the probability of 0 deaths is $q_1 q_2 q_3$. The probability of one death is the probability of the first two patients surviving ($q_1 q_2$) times the probability that the third dies plus the probability that one of the first two dies ($q_1 p_2 + p_1 q_2$) times the probability that the third survives, or $(q_1 p_2 + p_1 q_2) \cdot p_3$. The probability of two deaths is the probability that one of the first two patients dies times the

probability that the third patient dies ($q_1p_2 + p_1q_2$)· p_3 plus the probability of the first two patients dying times the probability that the third survives $p_1p_2q_3$.

At each stage in the calculation, the probability of 0 deaths is merely the cumulative product of the q_i s. The probability of any other number of deaths is the probability of one less than that number (calculated for the previous number of patients) times the probability that the current patient dies plus the probability of that number of deaths among the previous number of patients times the probability that the last patient survives.

Following the explanation, the computational approach is straightforward. It is not necessary to store the whole array; one merely needs to keep track of the sums for the previous patient. Furthermore, while probabilities can be computed for all possible numbers of deaths, 0, 1 . . . n , in practice one is unlikely to be interested in using the exact calculation when the *expected* number of deaths is five or more. In most situations, if the expected number of deaths is five or less, then the likelihood of observing 15 deaths is vanishingly small. The program in Appendix B uses such a cutoff. This program also incorporates a running sum of the p_i s and p_iq_i . One could forgo further exact calculations when the Σp_i exceeds 5 and then rely on the normal approximation. Since the exact calculation is so efficient, however, the program uses the exact calculation in all instances in which the observed number of deaths is 15 or less.

APPENDIX B

SAS® Version of Calculations

```
* THIS PROGRAM USES AS INPUT A PATIENT-LEVEL DATASET, SORTED BY
HOSPITAL ID ("HOSPID"). IT HAS A VARIABLE CALLED "DIED", WHICH
IS A BINARY VARIABLE (1=YES, 0=NO) AND A VARIABLE CALLED
"P", WHICH CONTAINS THE PREDICTED MORTALITY SCORE.
IT CALCULATES FOR EACH HOSPITAL 1) PL, THE LOWER TAIL PROBABILITY OF
OBSERVING D DEATHS OR FEWER AND 2) PU, THE UPPER TAIL PROBABILITY OF
OBSERVING D DEATHS OR MORE.
IF THERE ARE 15 OR FEWER DEATHS IN A GIVEN HOSPITAL, IT CALCULATES
THE EXACT PROBABILITY. OTHERWISE, IT CALCULATES THE PROBABILITY
USING A NORMAL APPROXIMATION.;
DATA TEMP;
  SET _____ (KEEP=HOSPID DIED P);          * FILL IN INPUT DATA SET NAME;
  BY HOSPID;
  RETAIN SUMDIED SUMPRED SUMPQ SUMNPATS LAST0-LAST15;
  IF DIED=. OR P=. THEN DELETE;
  * WHEN STARTING A NEW HOSPITAL, RESET THE RUNNING TOTALS;
  IF FIRST.HOSPID THEN DO;
    SUMDIED = 0;
    SUMPRED = 0;
    SUMPQ = 0;
    SUMNPATS = 0;
  END;
  * INCREMENT THE RUNNING TOTALS;
```

APPENDIX B (Continued)

```

SUMDIED = SUMDIED + DIED;
SUMPRED = SUMPRED + P;
SUMNPATS = SUMNPATS + 1;
Q = 1 - P;
SUMPQ = SUMPQ + (P*Q);
* ONLY DO THE EXACT PROBABILITY CALCULATIONS WHILE THERE ARE 15 OR;
* FEWER DEATHS IN THIS HOSPITAL;
IF SUMDIED <= 15 THEN DO;
  ARRAY CURRENT (*) PROB0 - PROB15;
  ARRAY LAST      (*) LAST0 - LAST15;
  IF FIRST.HOSPID THEN DO;
    CURRENT{1} = Q;
    CURRENT{2} = P;
  END;
  ELSE DO;
    CURRENT{1} = LAST{1} * Q;
    IF SUMNPATS <= 15 THEN DO;
      DO J = 2 TO SUMNPATS;
        I = J - 1;
        CURRENT{J} = (LAST{I} * P) + (LAST{J} * Q);
      END;
      * THE VALUE OF J IS NOW SUMNPATS + 1;
      CURRENT{J} = LAST{J-1} * P;
    END;
    ELSE DO J = 2 TO 16;
      I = J - 1;
      CURRENT{J} = (LAST{I} * P) + (LAST{J} * Q);
    END;
  END;
  DO I = 1 TO 16;
    LAST{I} = CURRENT{I};
  END;
END;
IF LAST.HOSPID THEN DO;
  IF SUMDIED > 15 THEN DO;          * NORMAL APPROXIMATION;
    PL = PROBNORM((SUMDIED + 0.5 - SUMPRED) / SQRT(SUMPQ));
    PU = 1 - PROBNORM((SUMDIED - 0.5 - SUMPRED) / SQRT(SUMPQ));
  END;
  ELSE DO;                          * EXACT PROBABILITY;
    IF SUMDIED = 0 THEN DO;
      PU = 1;
      PL = CURRENT{1};
    END;
    ELSE DO;
      SUMPROBS = 0;
      INDX = INT(SUMDIED + 0.0001);
      DO I = 1 TO INDX;
        SUMPROBS = SUMPROBS + CURRENT{I};
      END;
      PU = 1 - SUMPROBS;
      PL = SUMPROBS + CURRENT{INDX+1};
    END;
  END;
  OUTPUT;
END;
KEEP HOSPID SUMNPATS SUMDIED SUMPRED PL PU;
PROC PRINT;

```

ACKNOWLEDGMENTS

We are grateful to Mark Blumberg, M.D., members of the Pew/AHCPR Writing Seminar at University of California-San Francisco, and anonymous reviewers for helpful comments on an earlier draft; to Dr. John Tukey for a suggestion outlining the method of computing the exact probabilities; to Deborah Peltzman Rennie for the SAS® program; and to Lucy Marton for preparation of the manuscript.

NOTES

1. Macintosh® is a registered trademark of the Apple Computer Corporation.
2. Microsoft® is a registered trademark of the Microsoft Corporation.
3. SAS® is a registered trademark of SAS Institute Inc.

REFERENCES

- Escarce, J. J., and M. A. Kelley. "Admission Source to the Medical Intensive Care Unit Predicts Hospital Death Independent of APACHE II Score." *Journal of the American Medical Association* 264, no. 18 (14 November 1990): 2389-94.
- Fleiss, J. L. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley and Sons, 1981.
- Flood, A. B., W. R. Scott, W. Ewy, W. H. Forrest, Jr., Byron W. Brown. "Tests for Differences among Hospitals in the Quality of Surgical Outcomes and Service Intensity." In *Hospital Structure and Performance*. Edited by A. Barry Flood, and W. R. Scott. Baltimore: The Johns Hopkins University Press, 1987.
- Knuth, D. E. *The Art of Computer Programming*. 2nd ed. Menlo Park, CA: Addison-Wesley Publishing, 1981.
- Murphy, E. A. *Probability in Medicine*. Baltimore: The Johns Hopkins University Press, 1979.
- Weiss, K. B., and D. K. Wagener. "Changing Patterns of Asthma Mortality: Identifying Target Populations at High Risk." *Journal of the American Medical Association* 264, no. 13 (3 October 1990): 1683-87.